NASA's NPOESS Preparatory Project Science Data Segment: A Framework for Measurement-Based Earth Science Data Systems

Mathew R. Schwaller and Robert J. Schweiss
NPP Project, NASA/Goddard Space Flight Center, Greenbelt, MD, USA
email: mathew.r.schwaller@nasa.gov

***Abstract*— The NPOESS Preparatory Project (NPP) Science Data Segment (SDS) provides a framework for the future of NASA's distributed Earth science data systems. The NPP SDS performs research and data product assessment while using a fully distributed architecture. The components of this architecture are organized around key environmental data disciplines: land, ocean, ozone, atmospheric sounding, and atmospheric composition. The SDS thus establishes a set of concepts and a working prototypes. This paper describes the framework used by the NPP Project as it enabled Measurement-Based Earth Science Data Systems for the assessment of NPP products.**

Keywords: NPP, NPOESS, Earth science, data systems, SDS

I. INTRODUCTION

The NPOESS Preparatory Project (NPP) Science Data Segment (SDS) provides a framework for the future of NASA's distributed Earth science data systems. The NPP SDS performs research and data product assessment while using a fully distributed architecture. The components of this architecture are organized around key environmental data disciplines: land, ocean, ozone, atmospheric sounding, and atmospheric composition. The SDS thus establishes a set of concepts and working prototypes for generating Climate Data Records (CDRs) focused on key Earth science themes: Atmospheric Composition, Climate Change, Carbon/Ecosystems, Solid Earth, Weather, and Water/Energy Cycle. The NPP SDS is a measurement-based system in the sense that its focus is on CDRs, rather than mission-unique data products.

The NPP SDS is one of NASA's contributions to risk reduction for the National Polar-orbiting Operational Environmental Satellite System (NPOESS). NPOESS is currently in development as the U.S. Government's future low-Earth orbiting satellite system for monitoring global weather and the environment. NPOESS is a joint mission of three federal agencies: the National Aeronautics and Space Administration (NASA), the Department of Defense (DoD), and the National Oceanic and Atmospheric Administration (NOAA). The NPOESS Preparatory Project (NPP) provides a "bridge" between NASA's current Terra, Aqua, and Aura missions and future NPOESS missions. NPP also provides a pre-operational, on-orbit test and risk reduction for key NPOESS instruments and ground-based data processing capabilities while maintaining continuity of environmental data used for long-term climate change research.

The NPP Mission is planning for a launch readiness date in September 2009. NPP will have an 824 km polar, sun-synchronous orbit with a 10:30 AM descending[1] node crossing. The NPP mission provides remotely sensed land, ocean, atmospheric, ozone, and sounder data that will serve the meteorological and climate communities. The NPP SDS completed a system requirements review in April 2005 and a preliminary design review in Sept 2006. A NPP SDS Critical Design Review is planned for October 2007, with additional milestones leading to full research-operations in time for the currently planned, September 2009, NPP launch date.

II. SDS OBJECTIVES

The primary purpose of the NPP SDS is to "assess the quality of the NPP EDRs for accomplishing climate research" [1]. These Environmental Data Records (EDRs) are part of the suite of NPP/NPOESS data products that also include Raw Data Records (RDRs), Sensor Data Records (SDRs), and a SDR variant for microwave sensors called a Temperature Data Record (TDR). RDRs are processed to eliminate communications artifacts only; SDRs remove the sensor signature and apply calibration to recreate the flux received at the sensor aperture; and EDRs employ environmental models to estimate biogeophysical parameters of interest to the Earth science community [2].

Four NPP sensors will be flown on the NPP observatory: the Visible Infrared Imaging Radiometer Suite (VIIRS), which provides global observations of land, ocean, and atmospheric parameters; the Cross-Track Infrared Sounder (CrIS), which provides cross-track measurements of scene radiances to permit the calculation of vertical distribution of temperature, moisture, and pressure in the earth's atmosphere; the Advanced Technology Microwave Sounder (ATMS), which collects specialized data to permit, in conjunction with CrIS, the calculation of the vertical moisture, temperature, and pressure profiles of the earth's atmosphere; and the Ozone Mapper/Profiler suite (OMPS), which uses three primary

[1] At the time of writing this paper the NPP/NPOESS Program is planning to change the orbit to a 1:30 PM Ascending.

sensors within a single instrument suite to perform complementary functions for atmospheric ozone monitoring.

There is one RDR for the VIIRS, CrIS, and ATMS sensors and three for the OMPS sensor, one SDR for each sensor or each sensor channel, and (for NPP) one EDR for each of 24 specified environmental products for the 4 sensors described above. The responsibility for evaluation of these 24 EDRs is allocated to the SDS Product Evaluation And Test Environments (PEATEs), which are described in more detail in Section III of this document.

As noted above, the primary purpose of the NPP SDS is research and assess the quality of the NPP Environmental Data Records (EDRs) for accomplishing climate research. This objective is fundamentally different from that of past Earth science spaceflight missions. Previous NASA missions developed mission data systems to generate mission-specific data products. In the case of NPP, the goal is to assess the products and to ensure that they can be used to generate long-term measurements (CDRs) that may span several spaceflight missions. As such, NPP SDS was designed to be a prototype of a multi-mission *measurement-based system.* Section 3 of this document describes the NPP system elements that were designed to achieve this goal. Sections 4 and 5 define the characteristics of the SDS architecture and operations that are unique to measurement-based systems. Taken together, these characteristics provide a "recipe," or set of guidelines for systems like SDS that assess Earth science data as "climate quality"—or alternatively—that conduct the research needed to enable the generation of CDRs from spaceflight mission data.

III. SDS System Elements

The SDS PEATEs have the primary responsibility for the NPP product evaluation, and for making recommendations on algorithm enhancements or other improvements to these products. Each PEATE leverages off of an existing Earth science data system or a center of demonstrated Earth science expertise. Thus each PEATE joins the SDS with a mature set of resources, expertise, and knowledge of the science in their respective discipline, as described below.

Ocean PEATE leverages off of existing resources from the Seaviewing Wide Field-of-view Sensor (SeaWiFS) Ocean Data Processing System (ODPS). The ODPS is a science investigator-led processing system (SIPS) that processes, calibrates, validates, archives, and distributes data received from the SeaWiFS instrument aboard the OrbView-2 observatory.

Land PEATE leverages off of the Moderate Resolution Imaging Spectroradiometer (MODIS) Data Processing System (MODAPS), also a SIPS. MODAPS currently generates Level 2 through Level 4 MODIS science products for distribution to NASA's Distributed Active Archive Centers (DAACs) for archival and to the MODIS science team for quality control.

Ozone PEATE leverages off of the ozone data product system, OMIDAPS [3], an existing SIPS that processes data from the Ozone Monitoring Instrument (OMI) instrument aboard the Aura satellite to higher-level science data products.

Atmosphere PEATE leverages University of Wisconsin-Madison's Space Science and Engineering Center (SSEC) expertise which includes, among other things, routine validation of the GOES Imager and Sounder products.

Sounder PEATE leverages off of the Atmospheric Infrared Sounder (AIRS) Project Team Leader Science Computing Facility (TLSCF) located at NASA/Jet Propulsion Laboratory which currently ingests AIRS, AMSU and HSB Level 0 data from the GSFC DAAC and operationally producing Level 1B1 products.

The 5 PEATEs described above are core elements of SDS. In this capacity, PEATEs serve as a liaison to the NASA-funded NPP Science Team, from which they draw on expertise for analysis and algorithm improvement. The PEATEs also provide scientific expertise in their own right, and they also provide the computing power needed for assessment of NPP RDRs, SDRs, and EDRs.[2]

In the pre-launch time frame, each PEATE acquires, adapts, or wraps SDR and EDR product generation software from the NPP operational system, known as the Interface Data Processing Segment (IDPS). Each PEATE integrates the operational software into its own environment and evaluates the science algorithms as implemented in the operational IDPS. Additionally, the SDS PEATES develop, test, and demonstrate corrections and enhancements to the operational code that may lead to algorithm improvement recommendations.

In the post launch phase, nominally, each PEATE acquires selected EDRs and corresponding RDR, SDR, intermediate products, and ancillary data, as needed for retrospective processing, evaluation, and internal distribution to the NPP Science Team members.

There are 4 other SDS system elements that support the PEATEs and the process of NPP product evaluation and improvement. These system elements have been developed specifically for SDS, and are sketched briefly below.

SDS Data Delivery Depository Element (SD3E) acquires NPP RDRs, SDRs, EDRs and Intermediate Products (IPs) from NOAA's Comprehensive Large Array-data Stewardship System (CLASS), NPP Science Investigator-led Processing System (NSIPS), and from the NPP Interface Data Processing Segment (IDPS). It subscribes to the External Data Providers, and receives, ingests and stages data in a "rolling archive" for pick-up by the PEATEs and NICSE. SDS is capable of storing approximately 32 days of Science Data and 5 days of Intermediate Products.

NPP Instrument Calibration Support Element (NICSE) supports pre- and post-launch radiometric and geometric characterization and calibration assessment of the VIIRS instrument. More importantly the NICSE has the knowledge, skills, and tools to understand, assess, and recommend updates to the VIIRS instrument calibration.

---

[2] RDR = Raw Data Record ~ CEOS/NASA Level 1A
SDR = Sensor Data Record ~ CEOS/NASA Level 1B
EDR = Environmental Data Record ~ CEOS/NASA Level 2

Integration and Test System Element (I&TSE) is a scaled down version of the NPP Production IDPS that is used by other SDS elements (principally the PEATEs) for trouble shooting product anomalies, regenerating intermediate products, or demonstrating algorithm enhancements and calibration improvements.

Project Science Office Element (PSOE) provides management direction and science guidance to all SDS elements. PSOE also serves as the single point for NASA to relay results of EDR assessments and recommendations to the operational agencies on NPP algorithm and calibration updates or improvements.

IV. MEASUREMENT-BASED ARCHITECTURE

The SDS elements described above fit within a system architecture tailored to the unique requirements of measurement-based systems. Several aspects of this architecture have been described elsewhere [3] but the fundamental concepts are reiterated and expanded below with a particular emphasis on NPP SDS.

Distributed. The SDS distributed architecture was not a system design goal in itself. Rather, it is a consequence of the principle requirement to organize the SDS by science focus and by the key biogeophysical measurements needed to evaluate EDRs. The SDS PEATEs were therefore selected on the basis of scientific expertise, and they are located in institutions that are centers of excellence for that expertise.

Scalable. The SDS is scalable on several levels. Individual PEATEs are scalable in the sense that their processing systems have typically been built to take advantage of open systems hardware and software. Thus the Ozone PEATE data processing system, OMIDAPS, for example, can operate on a single host up to any number of individual or clustered commodity Unix workstations [4, 5]. The SDS is also scalable in the sense that there is no arbitrary limit to the number of PEATEs that may be part of the SDS. PEATEs may either be added or removed without affecting the overall SDS operations.

One of the SDS elements that ensures PEATE scalability is the SD3E. As noted above, the SD3E operates as a central source for all PEATEs to gain access to IDPS and CLASS data records. Without SD3E, the SDS requests for IDPS and CLASS would grow linearly with the number of PEATEs. Several PEATEs may request the same data product, but the SD3E will request it only once from the IDPS or CLASS and stage it for pick-up by all the PEATEs that need it. Even with this approach, the fully-loaded demand on IDPS/CLASS is expected to reach 4 Terabytes/day; but without SD3E the SDS demand for IDPS/CLASS products would be considerably higher.

Flexible. The suite of spaceflight missions and the computing infrastructure for utilizing spaceflight data will likely undergo considerable and relatively rapid change over the next several years. These changes will be led by new and unanticipated policy directives, by accelerating technological innovation, and by changes in science objectives. The consequences of such changes on SDS may include the addition or deletion of PEATEs, changes to the scope and number of instruments and mission that the PEATEs must support, changes in data sources, and changes to the number and scope of EDRs that must be evaluated. The SDS design is adaptable to such change in management policy and science objectives.

Extensible. The SDS elements are designed to be expanded, re-using Science Investigator-led Processing architectures and concepts whereby shifting design and development efforts from infrastructure to science algorithm integration and testing. The SIPS architecture is platform independent and merely requires additional computers to expand processing throughput.

Science Data Stewardship. The SDS has played a key role in the responsibility of assuring the data is seamless and consistent over missions, science algorithms are documented sufficiently well, data provenance is adhered to [7], science data formats are clear, complete, unambiguous, and platform and application independent. Additionally, the SDS supported proxy data generation from other instrument data sets, and plays an active role in pre launch and post launch calibration / validation program. Finally, it has been the experience of the NPP SDS team that long term archiving requires special considerations including: complete capture of software used to produce science products, associated documentation, and golden data sets to demonstrate the software.

Configuration Management. All science algorithms, documents, and data, received, derived, modified, and used to evaluate and assess mission products must be carefully preserved. Ability to recreate diagnostic data records requires all of the original software and inputs including calibration coefficients, auxiliary data sets, look up tables and intermediate products. This necessitates a large-scale electronic library with capability of sharing information among many users in a distributed environment.

V. MEASUREMENT-BASED OPERATIONS

Resident expertise. A key aspect of the SDS design is that the data system is built on core capabilities within the science community. The decision to draw on this expertise greatly reduced the cost and risk to SDS development. SDS was able to exploit the significant investment that NASA had already made in Science Investigator-led Processing Systems (SIPS), for example, which formed the basis for several of the PEATEs. All of the PEATEs have been peer-reviewed in their areas of expertise. Indeed, many of the NPP EDR algorithms are based on processing code that originated in the home institutions of the SDS PEATEs. In this respect, all PEATEs have unique and well-demonstrated capabilities for EDR evaluation and enhancement.

Liaison to science teams. The success of SDS in evaluating EDRs relies on close cooperation with NPP Science Team members. The NPP Science Team includes experts familiar with the development of algorithms to retrieve Earth system properties from sensors such as VIIRS, CrIS, ATMS and OMPS to study global climate change. Some Team members are actively involved in research that links sensor physics to EDRs. Other Team members represent specific

Earth science disciplines and can evaluate EDRs in terms of the adequacy as climate-quality products. If the products are found to be inadequate for this purpose, Science Team members work with the PEATEs to propose adding ancillary data, revised calibration, or additional processing steps as necessary to bring EDRs up to climate quality. A close interaction between the NPP Science Team members and the PEATEs is expected. The process of product evaluation, diagnosis and enhancement is expected to continue throughout the nominal 5 year NPP mission after launch.

Characterization and Validation. Long-term radiometric and geometric stability are essential characteristics of climate-quality data. The SDS therefore places a great deal of emphasis on ensuring temporal continuity of the NPP EDRs and in ensuring that the NPP products are inter-comparable with similar data products from other Earth science spaceflight missions. In the case of the Ocean PEATE, for example, the initial focus is on validating that EDRs are based on a dataset where top-of-the-atmosphere radiances are consistent in time and space. Additional objectives include ensuring that the NPP data are sufficiently well characterized such that they can be part of a multi-mission data set. To ensure this calibration continuity, the Ocean PEATE has developed a set of tools including 1) a comprehensive bio-optical database, 2) programs to evaluate different atmospheric correction algorithms, 3) programs to link the calibration of different satellite instruments, 4) programs (including sensor cross-calibration) to develop a consistent *in situ* calibration data set, and 5) a set of methods to combine instrument measurements into a single time series [5]. Other PEATEs have similar methods to verify that the within-instrument and between-instrument calibration stability are suitable for the generation of long-term climate records.

Retrospective Processing. Nearly all the SDS PEATEs will have the capability to run EDR and SDR algorithms, i.e., adapted production or heritage algorithms, within their own computing environments. This capability is tuned to generating research or diagnostic data products for the purpose of characterization and EDR assessment. That is, products may be limited in temporal duration and spatial extent for evaluation of a specific phenomenon. The PEATEs will also have the capacity to generate global products for long time series, and to retrospectively process these products while adjusting calibration parameters or algorithms.

Algorithm improvement. While the goal of SDS is NPP data product evaluation, the ultimate result of such evaluation will be an ongoing set of recommendations for product improvement. Such improvements may take the form of geometric and radiometric calibration table updates. Improvements may also take the form of revised algorithms. As a whole, the SDS will have the ability to test and demonstrate improvements in the I&TSE. Demonstrating improvements in the I&TSE environment ensures that the improvements are not artifacts of the PEATE computing environment and that they are scientifically valid improvements reviewed and approved by the Science Team members. Additionally, demonstrating the improvements in the IDPS environment is intended to simplify and expedite the research to operations transition of algorithm changes.

The PSOE is responsible for communicating algorithm improvements from the SDS PEATES to the organizations within NPP/NPOESS program that are responsible for day-to-day operations and data production. After EDR improvements are vetted within the PEATEs and the NPP Science Team, any recommendations for product improvement are passed to the PSOE. The PSOE further evaluates the recommendations in terms of their scientific and programmatic significance. It is the PSOE that has direct liaison with NOAA and NPP operations staff. The PSOE therefore manages the transfer of recommendations to NPP operations and verifies that recommendations are implemented correctly.

VI. Conclusions

Measurement-based data systems will play a key role in the future of Earth Science as NASA transitions toward the generation of multi-mission Climate Data Records. The NPP SDS provides a framework for a measurement-based system architecture and operations. In general, measurement-based Earth science data systems will be make use of a distributed network of existing institutions that have the knowledge, skills, and expertise needed to generate CDRs. In the case of the NPP SDS, a measurement-based framework will assess the value of NPP EDRs as climate-quality data products. In the event that the EDRs are found to be less than climate quality, the SDS will make recommendations on EDR improvements to the operational agencies involved in the NPP and NPOES program. In any case, the SDS measurement-based framework will help ensure that climate-quality data records are available to the Earth science research community.

References

[1] NPP Level 1 requirements, Oct 28, 2004,

[2] A.M. Goldberg. 2002. Environmental data production and delivery for NPOESS. International Geoscience and Remote Sensing Symposium,. 24-28 June 2002, pp 999-1001 vol. 2.

[3] E.F. Stocker. The GPM Precipitation Processing System. Pp 132-143 in Sensors, Systems, and Next-Generation Satellites VII, R. Meynat et al., eds., Proceedings of SPIE, vol. 5234.

[4] C. Tilmes, M. Linda, A.J. Fleig. 2004. Development of two Science Investigator-led Processing Systems (SIPS) for NASA's Earth Observing System (EOS). Proceedings of the International Geoscience and Remote Sensing (IGARSS) Symposium 2004.

[5] R.A. Barnes, D.K. Clark, W.E. Esaias, G.S. Fargion, G.C. Feldman, C.R. McClain. 2003. Development of a consistent multi-sensor global ocean colour time series. International Journal of Remote Sensing. 24(20):4047-4064.

[6] R. Schweiss, R. Ullman, E. Ho. 2006 NPP SDS: A Brief Overview, Proceedings of the International Geoscience and Remote Sensing (IGARSS) Symposium 2006.

[7] A. Fleig, C. Tilmes, 2006, Provenance and Reuse: Essential Elements for Long-Term Climate Data Sets. American Geophysical Union Fall 2006 symposium, abstract IN13D-1180.